

A Distributed Algorithm for Demand Response with Mixed-Integer Variables

Sleiman Mhanna, *Student MIEEE*, Archie C. Chapman, *MIEEE* and Gregor Verbič, *Senior MIEEE*,

Abstract—This letter presents a distributed algorithm for aggregating a large number of households with mixed-integer variables and intricate couplings between devices. The proposed distributed gradient algorithm is applied to the double smoothed dual function of the adopted DR model. Numerical results show that, with minimal parameter adjustments, the convergence of the dual objective exhibits a very similar behavior irrespective system size.

Index Terms—Dual decomposition, accelerated gradient methods, demand response, smoothing technique, mixed-integer variables.

I. INTRODUCTION

EFFICIENT load scheduling and aggregation is a problem of growing importance in the area of demand response (DR). However, this problem is particularly difficult for two main reasons. First, many household electrical devices have discrete operating points that can only be represented by mixed-integer variables, and moreover, household device uses are often coupled, thus giving household electricity demand a combinatorial structure. Therefore, solving this problem centrally may spell intractability when the number of households is large. Second, solving this problem centrally would require sending all of the households' private information to the aggregator, which entails substantial communication overhead and privacy concerns.

Against this background, this letter proposes a distributed gradient algorithm applied to a double smoothed dual function. This work is not the first attempt to solve a DR problem with mixed-integer variables in a distributed fashion. This problem is decomposed in terms of devices and solved in a distributed fashion using the proximal bundle method in [1]. In contrast to [1], the DR problem in this work is decomposed in terms of households. Doing so, allows for a more expressive household model, which can incorporate the intricate couplings between storage devices, appliances and renewable energy resources.

II. DR MODEL AND PROBLEM DESCRIPTION

The model comprises a set of agents $\mathcal{I} := \{0, 1, 2, \dots, I\}$, where 0 is the aggregator and each $i \neq 0$ is a household agent. Let $x_i^t \in \mathbb{R}_+$ be the demand for electric energy for every agent $i \neq 0$ at time-slot $t \in \mathcal{T} := \{\tau, \dots, \tau+T-1\}$. The aggregator faces a set of cost functions $C^t : \mathbb{R}_+ \mapsto \mathbb{R}_+$, where $C^t(x_0^t)$ is the cost of supplying x_0^t units of energy to the households at time-slot t .

Given the households' feasible schedule sets $X_{i \neq 0}$ and their demand profile $\mathbf{x}_i = [x_i^\tau, \dots, x_i^{\tau+T-1}]$, the aggregator

can (centrally) minimise the total energy cost per scheduling horizon T by solving the following problem:

$$\min_{\mathbf{x}_i \in X_i} \sum_{t \in \mathcal{T}} C^t(x_0^t), \quad (1a)$$

$$\text{subject to} \quad \sum_{i \in \mathcal{I} \setminus 0} x_i^t = x_0^t, \quad t \in \mathcal{T}. \quad (1b)$$

The local constraints of agents $i \neq 0$ arise from the operating modes of different flexible loads including *interruptible* (e.g. PHEVs, pool pumps) and *non-interruptible* (e.g. washing machines, dishwashers) loads (as in [1], [2]). Problem (1) can also be written as $\mathcal{P}^* = \inf_{\mathbf{x} \in X} \{C(\mathbf{x}) : A_c \mathbf{x} = \mathbf{0}\}$, where $\mathbf{x} = \{\mathbf{x}_i\}_{i \in \mathcal{I}}$, $X = \prod_{i \in \mathcal{I}} X_i$, and A_c is the coupling constraint matrix.

Problem (1) is a mixed-integer program (MIP) that belongs to the class of NP-hard problems which are notorious for tending to be intractable (if solved centrally) when they grow in size. However, relaxing the coupling constraints (1b), through the Lagrangian relaxation method, bestows a separable structure on problem (1). The problem can then be decomposed into $I + 1$ independent subproblems that can be solved in parallel. The partial Lagrangian of problem (1) can be written as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{t \in \mathcal{T}} C^t(x_0^t) + \sum_{t \in \mathcal{T}} \lambda^t \left(\sum_{i \in \mathcal{I} \setminus 0} x_i^t - x_0^t \right), \quad (2)$$

where $\boldsymbol{\lambda} = [\lambda^\tau, \dots, \lambda^{\tau+T-1}]$ is the vector of Lagrange multipliers. Therefore, the Lagrange dual function would be

$$\mathcal{D}(\boldsymbol{\lambda}) = \inf_{\mathbf{x}_i \in X_i} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathcal{D}_0(\boldsymbol{\lambda}) + \sum_{i \in \mathcal{I} \setminus 0} \mathcal{D}_i(\boldsymbol{\lambda}), \quad \text{where} \quad (3)$$

$$\mathcal{D}_0(\boldsymbol{\lambda}) = \inf_{\mathbf{x}_0 \in X_0} \sum_{t \in \mathcal{T}} (C^t(x_0^t) - \lambda^t x_0^t), \quad \text{and} \quad (4)$$

$$\mathcal{D}_i(\boldsymbol{\lambda}) = \inf_{\mathbf{x}_i \in X_i} \sum_{t \in \mathcal{T}} \lambda^t x_i^t, \quad i \in \mathcal{I} \setminus 0. \quad (5)$$

Finally, the dual problem is $\max_{\boldsymbol{\lambda} \geq \mathbf{0}} \mathcal{D}(\boldsymbol{\lambda})$. However, in this DR scenario, the concave dual function $\mathcal{D}(\boldsymbol{\lambda})$ is typically nondifferentiable. Specifically, as the subproblems in (5) can have multiple optimal solutions for a given vector $\boldsymbol{\lambda}$, the dual function $\mathcal{D}(\boldsymbol{\lambda})$ can be nonsmooth. Consequently, applying a conventional gradient method [3] to this problem would exhibit a very slow convergence.

III. DOUBLE SMOOTHING METHOD

One way to obtain a smooth approximation of $\mathcal{D}(\boldsymbol{\lambda})$ is to modify the subproblems in (5) to ensure a unique optimal

solution for every λ . The dual function is modified as follows:

$$\mathcal{D}_\mu(\lambda) = \mathcal{D}_0(\lambda) + \sum_{i \in \mathcal{I} \setminus 0} \mathcal{D}_{i,\mu}(\lambda), \text{ where} \quad (6)$$

$$\mathcal{D}_{i,\mu}(\lambda) = \inf_{\mathbf{x}_i \in X_i} \left(\sum_{t \in \mathcal{T}} \lambda^t x_i^t + \frac{\mu}{2} \|\mathbf{x}_i\|^2 \right), \quad i \in \mathcal{I} \setminus 0, \quad (7)$$

and $\mu > 0$ is a smoothness parameter. The modified dual function $\mathcal{D}_\mu(\lambda)$ is smooth and its gradient $\nabla \mathcal{D}_\mu(\lambda) = A_c \mathbf{x}_{\mu,\lambda}$, where $\mathbf{x}_{\mu,\lambda}$ denotes the unique optimal solution of problem (6), is Lipschitz-continuous with Lipschitz constant $L_\mu = \frac{\|A_c\|^2}{\mu}$.

The aim of this smoothing is to obtain a Lipschitz-continuous gradient for which efficient smooth optimisation methods can be applied [4]. However, despite having a good convergence rate $\mathcal{D}_\mu(\lambda^*) - \mathcal{D}_\mu(\lambda_k)$ at iteration k when applying a fast gradient method, the same good rate of convergence does not apply to $\|\nabla \mathcal{D}_\mu(\lambda_k)\|$.

Since the aim is not only to efficiently solve the dual problem but also to recover a nearly feasible solution to the primal [5], a second smoothing is applied to the dual function to make it strongly concave. The new dual function is written as

$$\mathcal{D}_{\mu,\kappa}(\lambda) = \mathcal{D}_0(\lambda) + \sum_{i \in \mathcal{I} \setminus 0} \mathcal{D}_{i,\mu}(\lambda) - \frac{\kappa}{2} \|\lambda\|^2, \quad (8)$$

which is strongly concave with parameter $\kappa > 0$, and whose gradient $\nabla \mathcal{D}_{\mu,\kappa}(\lambda) = A_c \mathbf{x}_{\mu,\lambda} - \kappa \lambda$ is Lipschitz-continuous with constant $L_{\mu,\kappa} = L_\mu + \kappa$.

IV. FAST GRADIENT ALGORITHM

The fast gradient method involves two multiplier updates, $\lambda_{k+1} = \hat{\lambda}_k + \frac{1}{L_{\mu^k,\kappa^k}} \nabla \mathcal{D}_{\mu^k,\kappa^k}(\hat{\lambda}_k)$, and $\hat{\lambda}_{k+1} = \lambda_{k+1} + \frac{\beta^k}{\left(\sqrt{L_{\mu^k,\kappa^k}^k} - \sqrt{\kappa^k}\right)} (\lambda_{k+1} - \lambda_k)$, where $\beta^k = \left(\sqrt{L_{\mu^k,\kappa^k}^k} - \sqrt{\kappa^k}\right) \left(\sqrt{L_{\mu^k,\kappa^k}^k} + \sqrt{\kappa^k}\right)^{-1}$.

The parameters of the algorithm are set as follows, $\mu^{k+1} = \alpha^{k+1}/D_X$, and $\kappa^{k+1} = e^{(\log(\kappa^{\maxiter}/\kappa^1)/\maxiter)} \kappa^k$, where $\alpha^{k+1} = e^{(\log(\alpha^{\maxiter}/\alpha^1)/\maxiter)} \alpha^k$, and \maxiter is the maximum number of iterations. The distributed algorithm is described in Algorithm 1.

Algorithm 1 : Distributed algorithm

Parameters: $\lambda_1 \succeq 0$, $\kappa^1 > 0$, $\kappa^{\maxiter} = 0.0001$, $\hat{\mu}^{\min} \in [0.0001, 0.005]$, $\maxiter \in \{500, 1000\}$.

- 1: **Initialisation:** Households compute $D_{X_i} = \min \left\{ \frac{1}{2} \|\mathbf{x}_i\|^2 : \mathbf{x}_i \in X_i \right\}$, and send it to the aggregator which computes $D_X = \sum_{i \in \mathcal{I} \setminus 0} D_{X_i}$ and sets $\mu^1 = \alpha^1/D_X$, $\hat{\mu}^1 = \mu^1$, $\lambda_1 = \lambda_1$, $J = 1$ and $k = 1$.
- 2: **while** $k \leq \maxiter$ **do**
- 3: Aggregator solves $\mathcal{D}_0(\hat{\lambda}_k)$ and broadcasts $\hat{\lambda}_k$ and $\hat{\mu}^k$ to the households which solve $\mathcal{D}_{i,\hat{\mu}^k}(\hat{\lambda}_k)$ and return $\mathbf{x}_{i,\hat{\mu}^k,\hat{\lambda}_k}$ to the aggregator.
- 4: Aggregator computes $\nabla \mathcal{D}_{\hat{\mu}^k,\kappa^k}(\hat{\lambda}_k)$ and the primal $\mathcal{P}_r^k = \sum_{t \in \mathcal{T}} C^t \left(\sum_{i \in \mathcal{I} \setminus 0} \mathbf{x}_{i,\hat{\mu}^k,\hat{\lambda}_k}^t \right)$.
- 5: Aggregator computes $L_{\mu^k,\kappa^k}^k = \frac{\|A_c\|^2}{\mu^k} + \kappa^k$ and updates λ_{k+1} and $\hat{\lambda}_{k+1}$.
- 6: Aggregator updates μ^{k+1} and κ^{k+1} and sets $\{\hat{\mu}^{k+1} = \mu^{k+1} : \hat{\mu}^{k+1} \geq \hat{\mu}^{\min}\}$.
- 7: $k \leftarrow k + 1$.
- 8: **end while**

Output: Aggregator finds the best recovered primal solution \mathcal{P}_r^J along with $\hat{\lambda}_J$, $\hat{\mu}^J$ and $\mathbf{x}_{\hat{\mu}^J,\hat{\lambda}_J}$ such that $J := \{k : \mathcal{P}_r^J = \min\{\{\mathcal{P}_r^k\}_{k \in \{1, \dots, \maxiter\}}\}\}$.

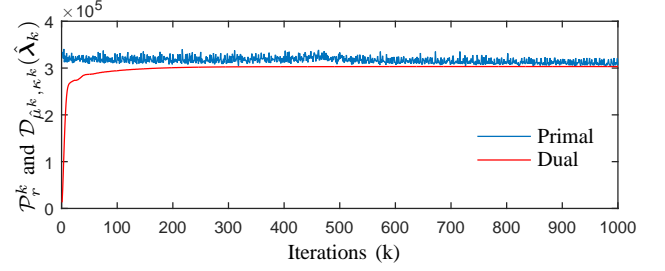


Fig. 1. Evolution of the primal and dual objectives for the $I = 1280$ case.

TABLE I
DIFFERENCE BETWEEN \mathcal{P}_r^J AND \mathcal{P}^* .

I	\mathcal{P}_r^J (\$)	\mathcal{P}^* (\$)	Gap (%)	$\hat{\mu}^{\min}$	κ^1
640	76049.78	75732.03	0.42	0.0004	10
1280	304089.52	302927.74	0.38	0.001	10
2560	1216367.92	1211711.03	0.38	0.0015	10

V. NUMERICAL EVALUATION

Algorithm 1 is tested on three scenarios, each with one aggregator and 640, 1280 and 2560 households respectively, each with 10 appliances scheduled over $T = 24$ h (as in [2]). In all three scenarios, Algorithm 1 is initialized with $\lambda_1 = 0$, $\kappa^1 = 10$, $\maxiter = 1000$, $\alpha^1 = 3 \times 10^{-4} \|A_c\|^2 D_X$ and $\alpha^{\maxiter} = 8 \times 10^{-8} \|A_c\|^2 D_X$. The simulation results along with the corresponding parameter values are listed in Table I. Table I shows that the difference between the recovered best integer feasible solution \mathcal{P}_r^J and the optimum solution \mathcal{P}^* does not exceed 0.42%, which corroborates the claim that a near-optimal solution can be recovered in a limited number of iterations. Finally, the evolution of \mathcal{P}_r^k and $\mathcal{D}_{\hat{\mu}^k,\kappa^k}(\hat{\lambda}_k)$ in the $I = 1280$ case is displayed in Figure 1, which also shows a quick and smooth convergence of the dual objective $\mathcal{D}_{\hat{\mu}^k,\kappa^k}(\hat{\lambda}_k)$ and a small duality gap upon termination.

VI. CONCLUSION

The aim of this work is to implement a fast gradient algorithm applied to the double smoothed dual function of a DR problem comprising expressive household models and mixed-integer variables. This work also demonstrates how to recover a near-optimal solution in a fixed number of iterations and minimal parameter tweaking.

REFERENCES

- [1] S.-J. Kim and G. Giannakis, "Scalable and robust demand response with mixed-integer constraints," *Smart Grid, IEEE Transactions on*, vol. 4, no. 4, pp. 2089–2099, Dec 2013.
- [2] S. Mhanna, G. Verbič, and A. Chapman, "A faithful distributed mechanism for demand response aggregation," *IEEE Trans. Smart Grid*, to be published.
- [3] B. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *lecture notes of EE392o, Stanford University, Autumn Quarter*, 2008.
- [4] Y. Nesterov, *Introductory lectures on convex optimization*. Springer, 2004, vol. 87.
- [5] O. Devolder, F. Glineur, and Y. Nesterov, "Double smoothing technique for large-scale linearly constrained convex optimization," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 702–727, 2012.